Capstone-1
Supplier Pricing Prediction and Segmentation

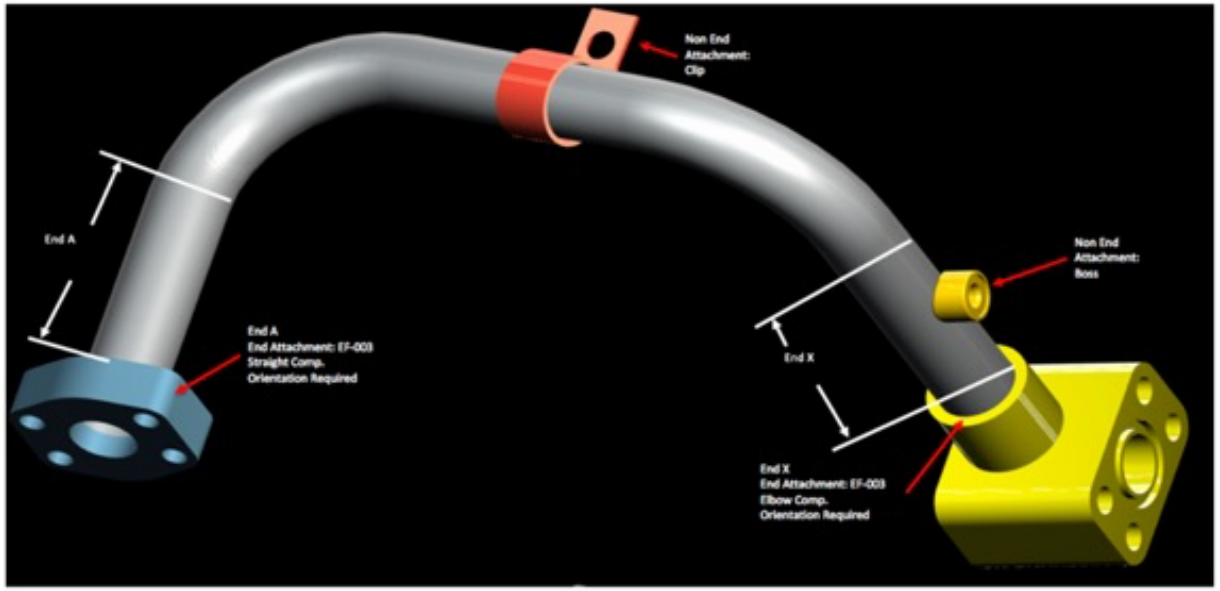
Domain: Supply Chain | Data Science

Motivation behind this Project?

Caterpillar (manufacturer of construction equipment) buys 8,855 unique tube assemblies of varying specifications from 57 different suppliers. Every time, company needs a tube assembly, they have to identify best qualified supplier and lowest best price, which has resulted in 30,213 requests for supplier quotes in 35 years.

Business Question:

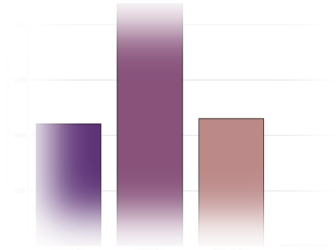
- Can we predict supplier pricing to save time and costs?
- Can we identify best managed assemblies and suppliers so that we can continuously improve?



About Assemblies: Which ones are Strategic?

4 Categories:

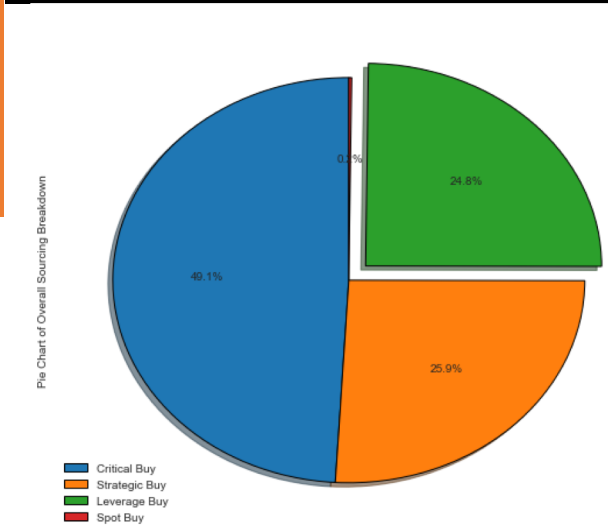
- Critical
- Strategic
- Leverage
- Spot

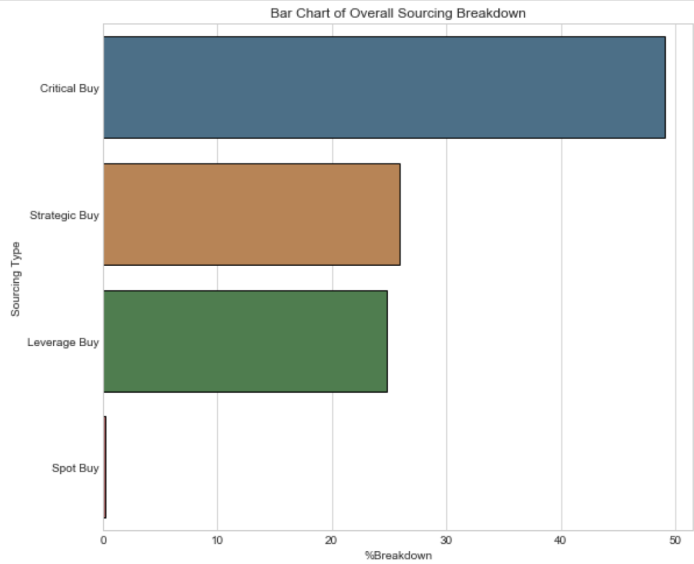


Years of High Spend : > 100 K

- 2005
- 2006
- 2009
- 2012
- 2013
- 2014

Assembly Category Breakdown:

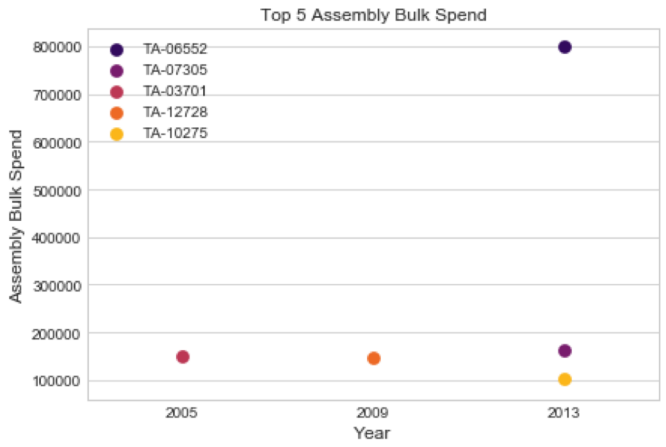


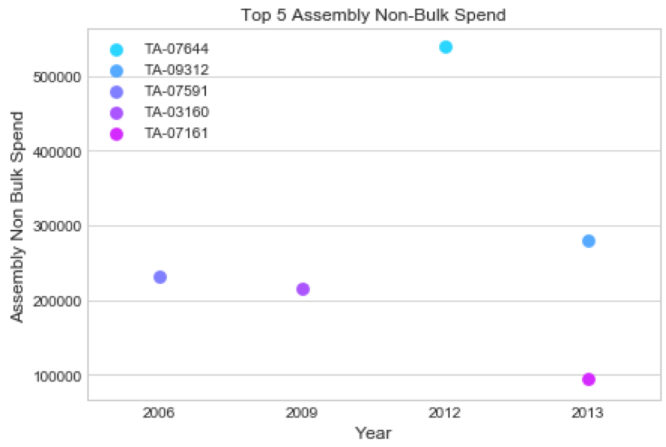


Observation:

1. Majority of tube assemblies are purchased from sole suppliers, 49.1%
2. Strategic and Leverage Options account 25.9% to 24.8% each.
3. There are opportunities to consolidate critical spend with leverage and strategic spends.

Top 5 Bulk and Non-Bulk Assembly Spends

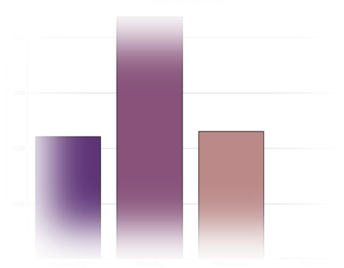




About Suppliers: Any Preferred Partners?

3Categories:

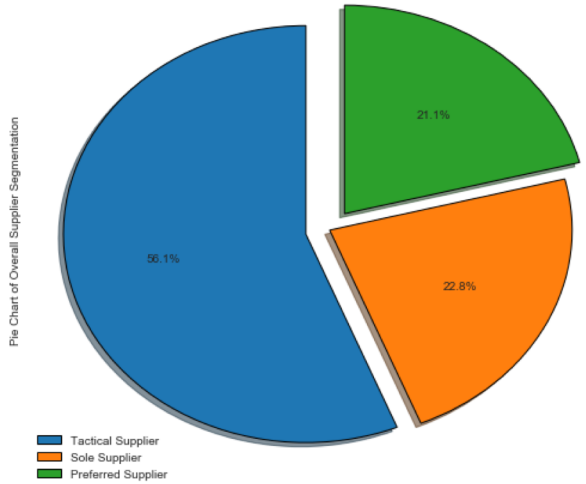
- Tactical
- Sole
- Preferred

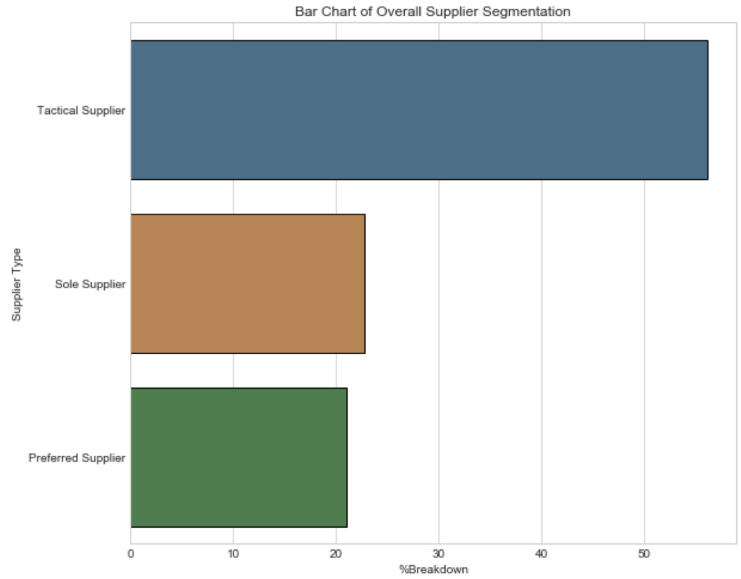


Pricing Types:

- Bulk Pricing
- Minimum Order Pricing

Supplier Segmentation:

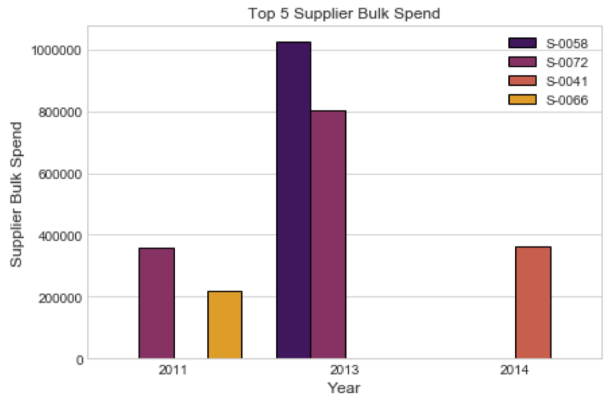


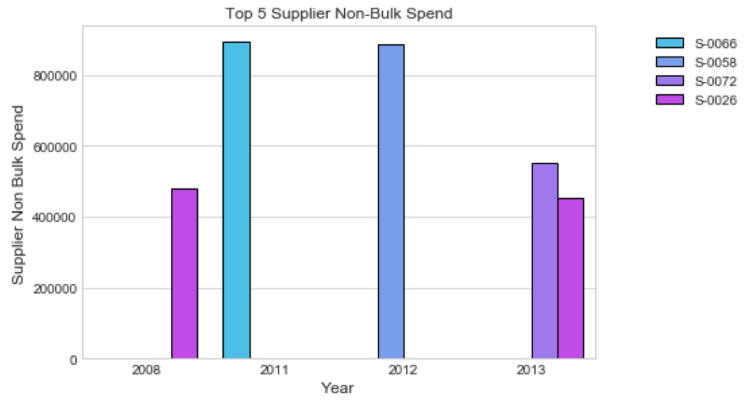


Observation:

1. High number of Tactical Suppliers, 56.1%
2. Fewer Preferred suppliers than Sole Suppliers.
3. There may be opportunities to transform tactical suppliers into Preferred.
4. May be an opportunity to conduct business review why we sole source?.

Top 5 Bulk and Non-Bulk Supplier Spends

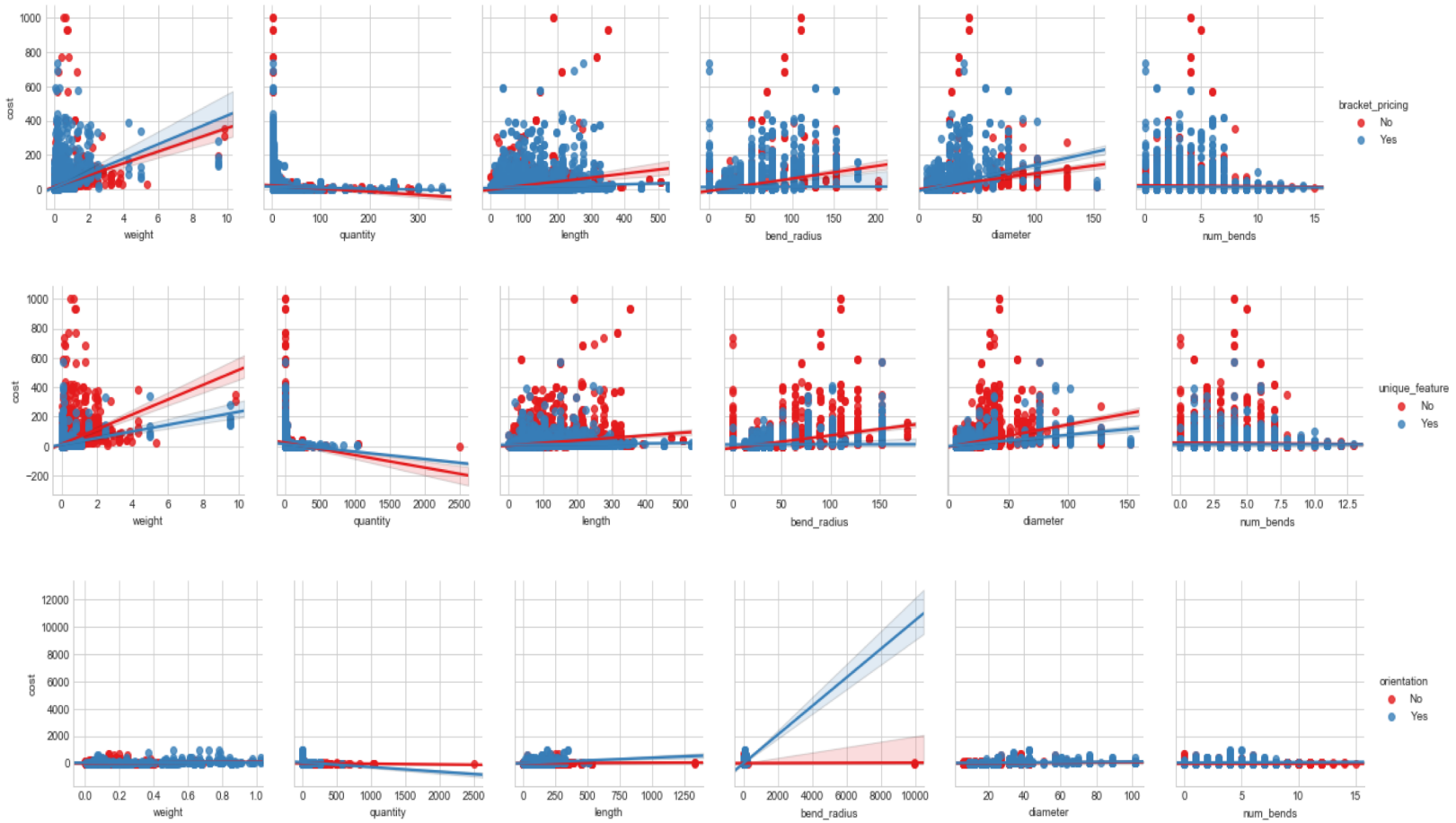




Input Parameters for Predicting Supplier Pricing

- Weight
- Length
- Diameter
- Bend Radius
- Number of Bends
- Orientation
- Pricing Type
- Unique Features
- Added Features
- Suppliers
- Components
- Material Specs
- Tube End Type
- Formed/Unformed Ends
- Quantity
- Usage
- Pricing Type

Dependency of Cost on Few Input Features



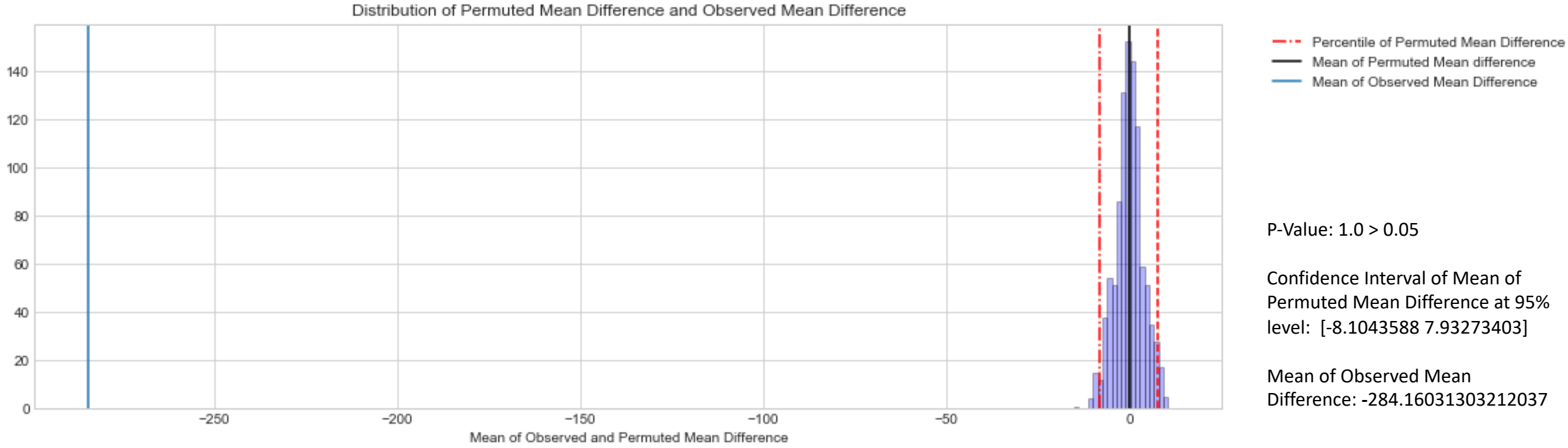
Observation:

1. Cost increases with increase in weight/length/bend radius/diameter and decreases with the increase in order quantity.
2. Cost decreases with increase in order quantity.
3. Heavier and bigger diameter bulk assemblies have higher cost increase.
4. Cost is not effected by the increase in number of bends.
5.. Assemblies with Unique features have lower cost increase. Could be due to lower demand.
6. Cost of having unique features is not impacted by increase in length, bend radius and number of bends.
7. Bulk assemblies with orientation have steep cost decrease.
8. Assemblies which are oriented and have Longer length and higher bend radius are more costly.
9. Orientation has negligible impact on assemblies weight and tube diameter.

Is Buying Bulk Assemblies on Minimum Order Basis a significant change?

Test Hypothesis by simulating 1000 samples as shown in the graph below.

- Null Hypothesis: Business will continue buying bulk assemblies in bulk.
- Alternate Hypothesis: Business will buy bulk assemblies as non-bulk (i.e. on minimum order basis)
- Calculate mean difference between observed and permuted samples for bulk and non-bulk assemblies.



P-Value: 1.0 > 0.05

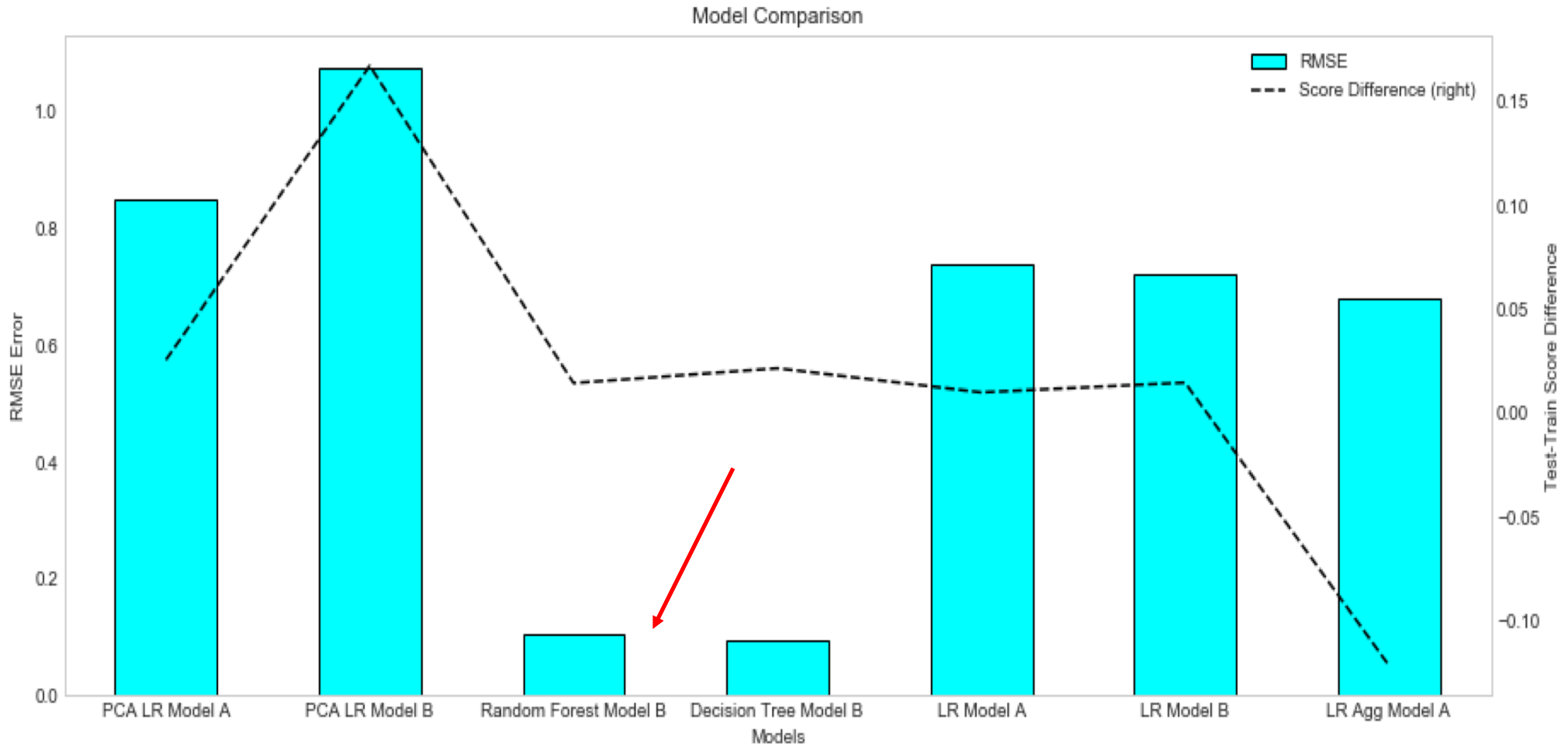
Confidence Interval of Mean of Permuted Mean Difference at 95% level: [-8.1043588 7.93273403]

Mean of Observed Mean Difference: -284.16031303212037

Observation:

1.High p-value value means that null hypothesis exists, even if we took 1000 samples to run the simulation.
2.Bulk assembly parameters will be different from the non-bulk assembly parameters, unless there is a significant shift in the observed mean difference.
3.Significant shift in the observed mean would mean that the bulk assembly parameters such as weight, costs and quantity have changed resulting in bulk assemblies being sourced as non-bulk (minimum order basis), which can be tracked and compared with the collected data.
4.When mean of observed mean difference becomes > 7.93, it lowers the p-value under 0.05, showing discontinuation in the business trend.

Supplier Price Prediction: Model Evaluation



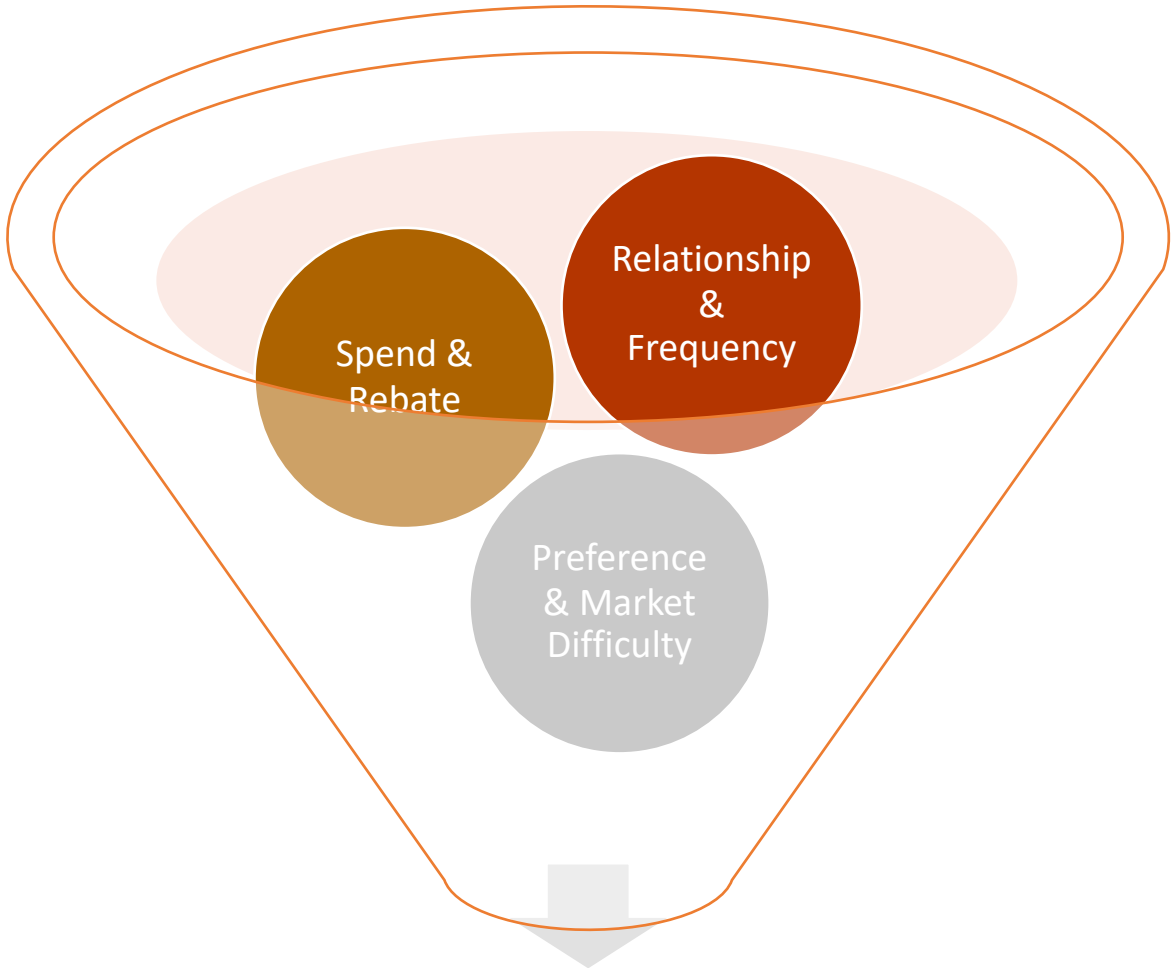
Observation:

1. Ensemble models resulted in lowest RMSE error as well as negligible difference between Train-Test score.
2. Random Forest is preferred over Decision Trees because it avoids model overfitting by employing multiple decision trees in the model.
3. **Supplier price was predicted with 98% accuracy on test validation data using Random Forest.**

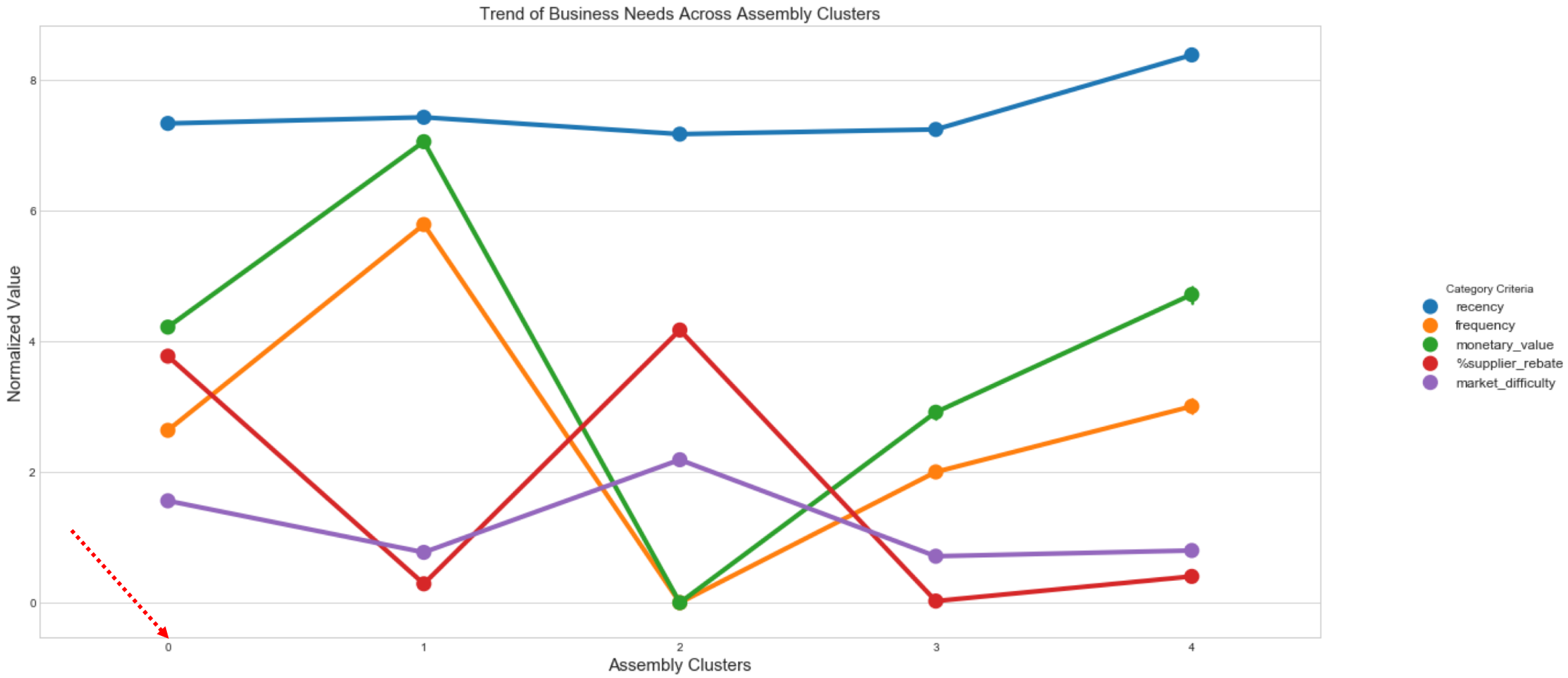
Criteria for Segmenting Assemblies and Suppliers

Business Needs:

- Spend: Total Spend Bulk/Non-Bulk
- Rebate: Discounts
- Relationship: Length of Contract
- Frequency: Number of Times Purchased
- Market Difficulty: Ease in switching supplier
- Preference: Supplier with maximum assortment
- Recency: Most Recent Purchase



Best Managed Assembly Cluster: 0

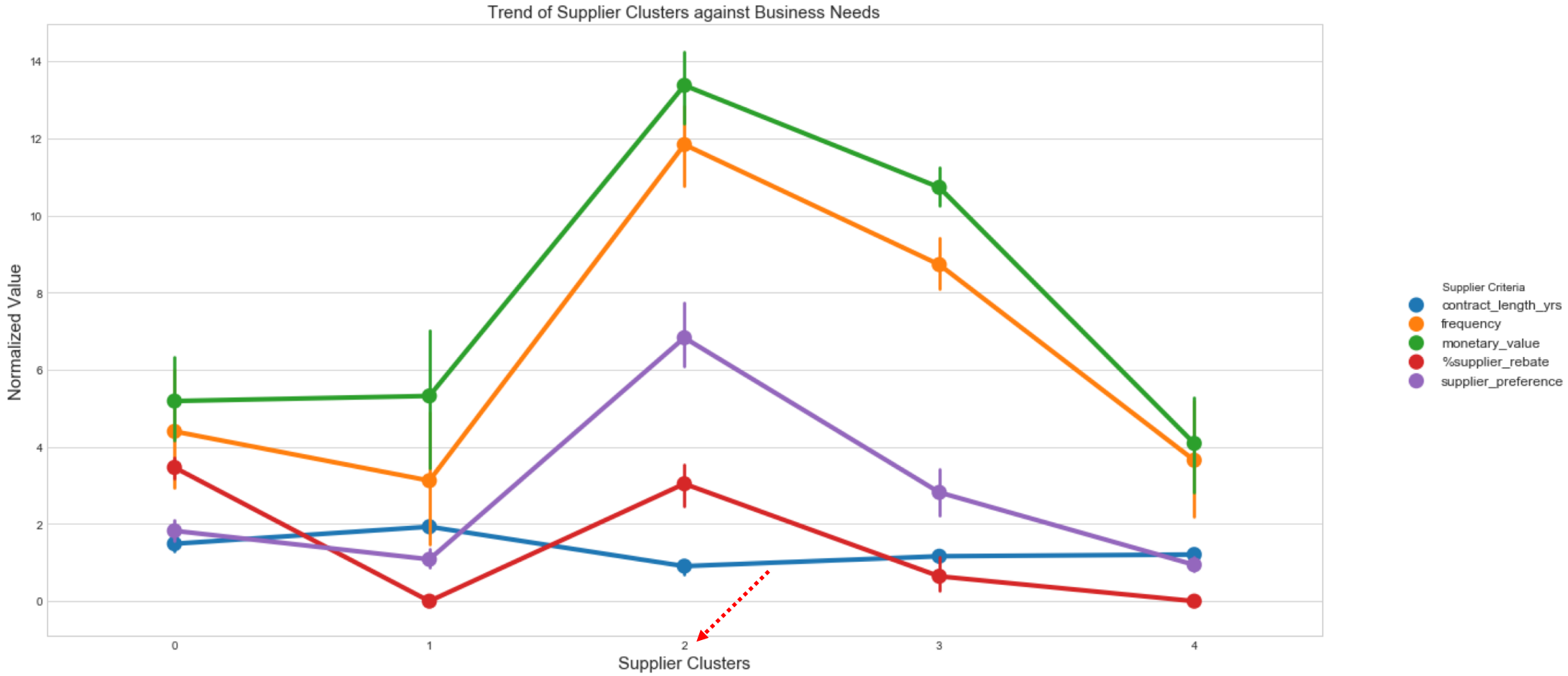


Observations:

1. Assembly Clusters 1,3 and 4 have high supplier spend and frequency of purchase but lowest supplier rebates.
2. Assembly Cluster 2 does not have supplier spend but are motivated to offer high rebates (potential for leverage buy).
3. Assembly Cluster 0 is the best managed category offering most rebates have fewer suppliers (like strategic buy).

Assembly Cluster 0 can be used as a benchmark to manage other clusters and find improvements.

Best Managed Supplier Cluster: 2



Observation:

1. Supplier clusters 1 and 4 has the lowest supplier preference and offer no rebates (similar to Sole Suppliers). Consider scope re-assignment/spend consolidation.
2. Supplier Cluster 2 is the best managed portfolio offering high rebates and have high supplier preference (Similar to Preferred Suppliers).
3. Supplier Cluster 0 demonstrates similar characteristics with cluster 2 but demonstrates very low supplier preference. There may be an opportunity to develop cluster 0 at par with cluster 2 and encourage healthy competition.
4. Supplier cluster 3 has shorter contract length, 2nd best supplier preference and significantly lower supplier rebates relative to high spend they manage. There should be an opportunity to negotiate better contract rates for a longer term deal.

Supplier Cluster 2 can be used as a benchmark to manage other suppliers and find improvements.

Project Conclusion

- Random Forest and K-means algorithm worked extremely well in predicting supplier pricing as well as categorizing assemblies and suppliers by business needs.

- As a result, we were able to predict how much an assembly of a given specification would cost to business, which category it would belong to and if the assembly would be supplied by our preferred supplier or not.

- This gave us the opportunity to establish benchmark out of a pool of 8,855 unique tube assemblies and 57 suppliers.

Future Possibilities

- Supplier Performance: We can add other features such as on-time delivery, safety statistics, contract compliance and inventory information to our existing model to gain further supplier insights.

- Real Time Dashboarding: We can connect our model to a real time data feed and monitor actionable insights using a dashboard.

- Further Modelling: We can try other algorithms such as time-series to predict annual supplier pricing.